

Amendments to the Claims:

This listing of claims will replace all prior versions, and listings of claims in the application:

Listing of Claims:

1. (Withdrawn) A computer implemented method for characterizing a plurality of biological sequences comprising:
obtaining a plurality of models, wherein each of the models represents a classification of biological sequences with structural or functional similarity; determining fitness of the biological sequences to the models; and automatically classifying the sequences according to the distances to the models.
2. (Withdrawn) The method of Claim 1 wherein the plurality of biological sequences have at least 50 sequences.
3. (Withdrawn) The method of Claim 2 wherein the plurality of biological sequences have at least 100 sequences.
4. (Withdrawn) The method of Claim 3 wherein the plurality of biological sequences have at least 100 sequences.
5. (Withdrawn) The method of Claim 3 wherein the models are Hidden markov models.
6. (Withdrawn) The method of Claim 5 wherein the classification is a family and each model represents a family.
7. (Withdrawn) The method of Claim 6 wherein the sequences are protein sequences.

8. (Withdrawn) The method of Claim 7 wherein the distances are E values.
9. (Withdrawn) The method of Claim 8 wherein the step of automatically determining comprises determining a step of determining a threshold for each of the models.
10. (Withdrawn) The method of Claim 9 wherein the step of determining a threshold comprises performing a curve analysis.
11. (Withdrawn) The method of Claim 10 wherein the step of performing a curve analysis comprises determining a point where the e-value curve drops abruptly or flattens.
- 12.-19. (Canceled)
20. (Withdrawn) A system for gene annotation comprising a processor; and a memory coupled with the processor, the memory storing a plurality of machine instructions that cause the processor to perform logical steps comprising obtaining a plurality of models, wherein each of the models represents a classification of biological sequences with structural or functional similarity; determining fitness of the biological sequences to the models; and automatically classifying the sequences according to the distances to the models.
21. (Withdrawn) The system of Claim 20 wherein the plurality of biological sequences have at least 50 sequences.
22. (Withdrawn) The system of Claim 21 wherein the plurality of biological sequences have at least 100 sequences.
23. (Withdrawn) The system of Claim 22 wherein the plurality of biological sequences have at least 100 sequences.
24. (Withdrawn) The system of Claim 23 wherein the models are Hidden markov models.

25. (Withdrawn) The system of Claim 24 wherein the classification is a family and each model represents a family.

26. (Withdrawn) The system of Claim 25 wherein the sequences are protein sequences.

27. (Withdrawn) The system of Claim 26 wherein the distances are E-values.

28. (Withdrawn) The system of Claim 27 wherein the step of automatically determining comprises determining a step of determining a threshold for each of the models.

29. (Withdrawn) The system of Claim 28 wherein the step of determining a threshold comprises performing a curve analysis.

30. (Withdrawn) The system of Claim 29 wherein the step of performing a curve analysis comprises determining a point where the e value curve drops abruptly or flattens.

31.-38. (Canceled)

39. (Withdrawn) A computer software product of the invention comprising a computer readable medium having computer-executable instructions for performing the method comprising:

obtaining a plurality of models, wherein each of the models represents a classification of biological sequences with structural or functional similarity;
determining fitness of the biological sequences to the models; and
automatically classifying the sequences according to the distances to the models.

40. (Withdrawn) The product of Claim 39 wherein the plurality of biological sequences have at least 50 sequences.

41. (Withdrawn) The product of Claim 40 wherein the plurality of biological sequences have at least 100 sequences.

42. (Withdrawn) The product of Claim 41 wherein the plurality of biological sequences have at least 100 sequences.

43. (Withdrawn) The product of Claim 42 wherein the models are Hidden markov models.

44. (Withdrawn) The product of Claim 43 wherein the classification is a family and each model represents a family.

45. (Withdrawn) The product of Claim 44 wherein the sequences are protein sequences.

46. (Withdrawn) The product of Claim 45 wherein the distances are E-values.

47. (Withdrawn) The product of Claim 46 wherein the step of automatically determining comprises determining a step of determining a threshold for each of the models.

48. (Withdrawn) The product of Claim 47 wherein the step of determining a threshold comprises performing a curve analysis.

49. (Withdrawn) The product of Claim 48 wherein the step of performing a curve analysis comprises determining a point where the e-value curve drops abruptly or flattens.

50.-57. (Canceled)

58. (New) A computer implemented method for gene characterization, the computer-implemented method comprising:

generating a plurality of libraries associated with a plurality of models related to a plurality of structural relationships of known proteins;

receiving a plurality of protein sequences;

processing information associated with the plurality of protein sequences and the plurality of models;

assigning the plurality of protein sequences to the plurality of models based on at least information associated with the plurality of protein sequences and the plurality of models.

59. (New) The method of claim 58 wherein the plurality of models are associated with a plurality of hidden markov models.

60. (New) The method of claim 58 wherein the plurality of protein sequences comprises 50 protein sequences.

61. (New) The method of claim 60 wherein the plurality of protein sequences comprises 150 protein sequences.

62. (New) The method of claim 61 wherein the plurality of protein sequences comprises 500 protein sequences.

63. (New) The method of claim 58 wherein the processing information associated with the plurality of protein sequences and the plurality of models comprises:
determining a plurality of thresholds associated with the plurality of models respectively based on at least information associated with the plurality of protein sequences and the plurality of models.

64. (New) The method of claim 63 wherein the determining a plurality of thresholds comprises processing information associated with a plurality of curves related to the plurality of models respectively.

65. (New) The method of claim 64 wherein the processing information associated with a plurality of curves related to the plurality of models respectively comprises processing information associated with a plurality of slopes related to the plurality of curves respectively.

66. (New) A system for gene annotation comprising a processor and a memory coupled with the processor, the memory storing a plurality of machine instructions that cause the processor to perform logical processes comprising:

generating a plurality of libraries associated with a plurality of models related to a plurality of structural relationships of known proteins;

receiving a plurality of protein sequences;

processing information associated with the plurality of protein sequences and the plurality of models;

assigning the plurality of protein sequences to the plurality of models based on at least information associated with the plurality of protein sequences and the plurality of models.

67. (New) The system of claim 66 wherein the plurality of models are associated with a plurality of hidden markov models.

68. (New) The system of claim 66 wherein the plurality of protein sequences comprises 50 protein sequences.

69. (New) The system of claim 68 wherein the plurality of protein sequences comprises 150 protein sequences.

70. (New) The system of claim 69 wherein the plurality of protein sequences comprises 500 protein sequences.

71. (New) The system of claim 66 wherein the processing information associated with the plurality of protein sequences and the plurality of models comprises:

determining a plurality of thresholds associated with the plurality of models respectively based on at least information associated with the plurality of protein sequences and the plurality of models.

72. (New) The system of claim 71 wherein the determining a plurality of thresholds comprises processing information associated with a plurality of curves related to the plurality of models respectively.

73. (New) The system of claim 72 wherein the processing information associated with a plurality of curves related to the plurality of models respectively comprises processing information associated with a plurality of slopes related to the plurality of curves respectively.

74. (New) A computer-readable medium including instructions for gene characterization, the computer-readable medium comprising:
one or more instructions for generating a plurality of libraries associated with a plurality of models related to a plurality of structural relationships of known proteins;
one or more instructions for receiving a plurality of protein sequences;
one or more instructions for processing information associated with the plurality of protein sequences and the plurality of models;
one or more instructions for assigning the plurality of protein sequences to the plurality of models based on at least information associated with the plurality of protein sequences and the plurality of models.

75. (New) The computer-readable medium of claim 74 wherein the plurality of models are associated with a plurality of hidden markov models.

76. (New) The computer-readable medium of claim 74 wherein the plurality of protein sequences comprises 50 protein sequences.

77. (New) The computer-readable medium of claim 76 wherein the plurality of protein sequences comprises 150 protein sequences.

78. (New) The computer-readable medium of claim 77 wherein the plurality of protein sequences comprises 500 protein sequences.

79. (New) The computer-readable medium of claim 74 wherein the one or more instructions for processing information associated with the plurality of protein sequences and the plurality of models comprise:

one or more instructions for determining a plurality of thresholds associated with the plurality of models respectively based on at least information associated with the plurality of protein sequences and the plurality of models.

80. (New) The computer-readable medium of claim 79 wherein the one or more instructions for determining a plurality of thresholds comprise one or more instructions for processing information associated with a plurality of curves related to the plurality of models respectively.

81. (New) The computer-readable medium of claim 80 wherein the one or more instructions for processing information associated with a plurality of curves related to the plurality of models respectively comprise one or more instructions for processing information associated with a plurality of slopes related to the plurality of curves respectively.

Appl. No. 09/683,264
Amdt. dated April 29, 2004
Reply to Office Action of December 30, 2003

PATENT

Amendments to the Drawings:

The attached sheet of drawings includes changes to Fig. 6. This sheet, which includes Fig. 6 replaces the original sheet including Fig. 6.

Attachment: Replacement Sheet
Annotated Sheet Showing Changes